A multi-scale priority model for smartphone actions

Jean-Pascal Pfister* and Arko Ghosh†

Jean-Pascal Pfister
Institute of Neuroinformatics and Neuroscience Center Zurich, University of Zurich / ETH Zurich

Arko Ghosh
Cognitive Psychology Unit, Institute of Psychology, Leiden University

(Dated: October 24, 2018)

Smartphone users touch the screen of the phone thousands of times per day. The inter-touch intervals follow a power-law distribution. We propose a multi-scale statistical model for those smartphone touches. At short-time scale, the model is governed by refractory effects, while at longer time scales, the touching rate is governed by the priority difference between smartphone tasks and other tasks. We show that both the statistics of the short intervals as well as the longer intervals are well captured by the model. The model is described in continuous time and the inter-touch interval distribution can be computed analytically.

Human actions such as mail correspondences, library loans or website visits are not equally distributed in time but are typically structured in bursts followed by long periods of inactivity [1, 2]. Priority-based models have been proposed to capture the power-law structure of the inter-event time distribution [3, 4]. More complicated models such as the cascading non-homogeneous Poisson process [5] provide a circadian explanation for the origin of power-law distributions. Here, we stick to the priority framework because of its simplicity and its tractability. We propose a generalisation of this priority-based model on different levels and apply it to smartphone touchscreen interaction data (see Fig. 1). First, our priority-based model does not only describe long inter-event intervals but also explicitly includes a relative refractory period after each event which helps to better describe short intervals and thereby overcomes the need to define an artificial onset of the power-law distribution [6]. Secondly, because our model is based on arbitrary priority distribution and not on specific priority distribution imposed by the presence of lists (with discrete number of items), it can produce any power-law exponent. Finally, our model is described in continuous time such that the inter-event interval distribution can be computed analytically which massively simplifies the fitting procedure. We found that for each subject, the inter-touch interval (ITI) distribution is different and well captured by the model. We also found that from those fitted parameters, we can quantify the priority placed on smartphone actions.

Discrete-time model. In the first step, we propose a discrete-time generative model for smartphone touches. This model extends existing priority-based models by including refractoriness [3, 4]. The output of the model is

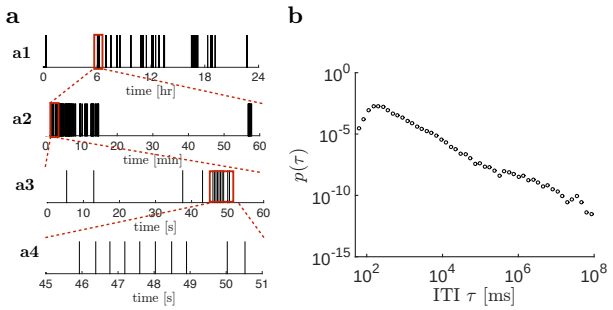

FIG. 1. Smartphone touch data. **a** Smartphone touch events (vertical bars) are characterised by bursts as well as long gaps at time scales of hours (**a1**) minutes (**a2**) and seconds (**a3**). At the sub-second time scale (**a4**), touches are more regular. **b** The inter-touch interval (ITI) distribution is scale free from seconds to hours. Data from one individual.

the set of touch times $\{t_0, t_1, \ldots, t_N\}$ where t_i can take discrete values, i.e. $t_i = k_i \Delta t$ with Δt being the bin width and $k_i \in \mathbb{N}$. Equivalently, the model output can be described by the touch train s_t where $s_t = 1$ denotes the presence of a touch while $s_t = 0$ indicates the absence of a touch.

Every touch is the result of a decision process. We assume that an individual can perform tasks from only two categories: either a task related to a smartphone screen *touch* or *other* task such as driving a car. In each category, there can be important tasks (such as dialling an emergency number) or less important tasks (such as checking the news). So we will assume that every task can be described by its priority level which is a number between 0 and 1. Let $x_t \in [0, 1]$ denote the priority associated with a *touch* task at time t and $y_t \in [0, 1]$ the priority associated to the *other* task. If at time t the *touch* task associated to priority x_t is executed (i.e. $s_t = 1$), then a new *touch* task is considered and will be attributed a new *touch* priority value drawn from the *touch* priority

* jpfister@ini.uzh.ch
† a.ghosh@fsw.leidenuniv.nl

distribution, i.e. $x_{\text{new}} \sim p(x)$. If the *touch* task is not executed ($s_t = 0$), the priority remains the same. This can be summarised as

$$x_{t+\Delta t} = x_t(1 - s_t) + x_{\text{new}} s_t \qquad x_{\text{new}} \sim p(x). \tag{1}$$

Conversely, the dynamics for the *other* priority y_t is such that when the screen is not touched at time t (i.e. $s_t = 0$), then it is the *other* action that is executed and a new priority y_{new} must be drawn from $q(y)$. This is summarised as

$$y_{t+\Delta t} = y_t s_t + y_{\text{new}}(1 - s_t) \qquad y_{\text{new}} \sim q(x). \tag{2}$$

In order to generate a smartphone touch, two conditions need to be satisfied. Firstly the priority x_t of the smartphone action needs to be greater than the priority y_t of the other action and secondly, the individual must be in a non-refractory state. Formally, the touch variable s_t is sampled from the following Bernoulli distribution:

$$s_t \sim \text{Bernoulli}(\lambda(x_t, y_t, \tau_t)\Delta t), \tag{3}$$

where the touching intensity λ (probability per time bin Δt) is given by

$$\lambda(x, y, \tau) = \rho r(\tau) H(x - y), \tag{4}$$

where $\tau = t - \hat{t}$ is the time since last touch ($\hat{t} = \max_{t_k}\{t_k < t\}$) and H is the Heaviside step function which guarantees that touches can only be generated when $x > y$ and ρ is the touching rate. $r(\tau) \geq 0$ is the refractory function which includes post-touch effects (i.e. right after a touch, the touch probability can be reduced). We express this refractory function as a sum of basis functions

$$r(\tau) = 1 + \sum_{k=1}^{n} \gamma_k \exp\left(-\alpha_k \tau\right), \tag{5}$$

with logarithmically spaced inverse time constants, i.e. $\alpha_k = \alpha_1 \beta^{-(k-1)}$. We took $\alpha_1^{-1} = 50$ ms, and set β such that $\alpha_8^{-1} = 1000$ ms. Note that the set $\{\gamma_k\}_{k=1}^n$ has to be chosen such that the condition $r(\tau) \geq 0$ is satisfied for all $\tau \geq 0$. The discrete-time model described by Eqs (1), (2) and (3) is a latent dynamical system. Note that sampling this model is slow since the complexity of this sampling scheme scales with the number of bins. Even more critical is the learning procedure for such a latent dynamical model which can be prohibitively slow for smartphone touching data sets which typically extend over months. A much faster sampling scheme is proposed below.

Continuous-time model. The idea of the continuous-time model is to directly sample the intervals τ instead of sampling the touch variable s_t at each time step. The transition to this continuous model can be done in two steps. First, we observe that when Δt is small, the *other* priorities y_t constantly change (except at the rare times where $s_t = 1$), i.e. Eq. (2) can be approximated as $y_t \sim q(y)$. This means that the priorities y_t are independent of time and therefore, the probability of generating a touch can be marginalised over y_t:

$$\begin{aligned} p(s_t|x_t, \tau_t) &= \int_0^1 p(s_t|x_t, y_t, \tau_t) q(y_t) dy_t \\ &= \text{Bernoulli}(\bar{\lambda}(x_t, \tau_t)\Delta t), \end{aligned} \tag{6}$$

where the average touching intensity $\bar{\lambda}$ is given by

$$\bar{\lambda}(x, \tau) = \int_0^1 \lambda(x, y, \tau) q(y) dy = \rho r(\tau) \pi(x), \tag{7}$$

and $\pi(x)$ is the probability of having $x > y$ for a given x

$$\pi(x) = \int_0^x q(y) dy. \tag{8}$$

In the second step, we take the limit $\Delta t \to 0$ and therefore, the inter-touch interval distribution conditioned on x can be expressed as (see also [7]):

$$p(\tau|x) = \bar{\lambda}(x, \tau) \mathrm{e}^{-\int_0^\tau \bar{\lambda}(x,t) dt}. \tag{9}$$

The unconditioned ITI distribution is obtained by averaging the conditioned ITI distribution over the *touch* priority distribution $p(x)$:

$$p(\tau) = \int_0^1 \bar{\lambda}(x, \tau) \mathrm{e}^{-\int_0^\tau \bar{\lambda}(x,t) dt} p(x) dx. \tag{10}$$

So samples of the continuous-time model can be simply obtained in a two-step procedure. First, x is sampled from $p(x)$, then τ is sampled from $p(\tau|x)$ given by Eq. (9). For this second step, one can use the time rescaling theorem [7]. Note that this continuous-time model describes a renewal process and hence the sampling complexity scales with the number of touches N.

Invariance of the model. Before giving a parametric form for all distributions, let us first note an invariant property of the model. In particular, it can be shown (see SI) that the ITI distribution given by Eq. (10) remains unchanged if the pair of priority distributions $(p(x), q(y))$ is replaced by $(\tilde{p}(x), \tilde{q}(y))$ given by

$$\tilde{p}(x) = p(\phi(x))\phi'(x) \qquad \text{and} \qquad \tilde{q}(y) = q(\phi(y))\phi'(y) \tag{11}$$

where ϕ is a differentiable and strictly monotonously increasing function with boundary conditions $\phi(0) = 0$ and $\phi(1) = 1$. This invariance can be understood intuitively by noting that the notion of priority contains some arbitrariness. Indeed, the only element which is relevant in the decision process is whether x is larger or smaller than y. If we define a new priority $x' = \phi(x)$ (with the above conditions on ϕ), we observe that the ordering remains unchanged, i.e. $x > y \Rightarrow \phi(x) > \phi(y)$. This observation can also be made more formally with a change of variable

in Eq. (10) (see SI). Secondly, this invariance property of the model means that without loss of generality, we can set one distribution and rescale the other one. For example, without loss of generality, we can set $q(y) = 1$. For the *touch* priority distribution, we will assume that it is given by a Beta distribution:

$$p(x) = \text{Beta}(x; a, b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \tag{12}$$

where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$ is the Beta function. With the above choice of q, the ITI distribution in Eq. (10) can be rewritten in a simpler form

$$p(\tau) = \rho r(\tau) \int_0^1 x \mathrm{e}^{-x\rho \int_0^\tau r(t)dt} p(x)dx \tag{13}$$

Scale free inter-touch interval distribution. For short time scales ($\tau < \tau_n$), the ITI distribution is governed by the refractory function r (see Fig. 2a). However, for longer time scales ($\tau \gg \tau_n$), the ITI distribution follows a power-law distribution. This can be seen in two steps. First, in the limit of large τ, we have $r(\tau) \to 1$. Secondly, in the limit of large τ, we know from Eq. (13) that the ITI distribution is only sensitive to the *touch* priority distribution in the vicinity of $x = 0$ that we denote as $p_0(x)$. Note that $p_0(x)$ is not normalised. For the Beta distribution, we have $p(x) \to p_0(x) = x^{a-1}/\text{B}(a,b)$ when $x \to 0$. Therefore, when $\tau \gg \tau_n$, the ITI distribution can be approximated as

$$\begin{aligned} p(\tau) &\simeq \rho \int_0^1 x p_0(x) \mathrm{e}^{-x\rho\tau} dx \\ &\simeq \frac{\Gamma(a+1)}{\text{B}(a,b)\rho^a} \tau^{-(a+1)}, \end{aligned} \tag{14}$$

where $\Gamma(z) = \int_0^\infty x^{z-1}\mathrm{e}^{-x}dx$ is the Gamma function. Therefore the power-law exponent is given by $a+1$ (see Fig. 2b)

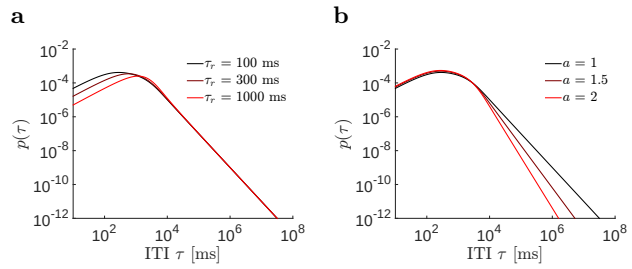

FIG. 2. Properties of the smartphone touching model. **a** The refractory time constant affects the early part of the ITI distribution. $n = 1$, $\tau_r = \alpha_1^{-1}$. **b** The parameter a from the priority distribution affects the power-law exponent of the ITI distribution.

List of models. For each subject, we fitted 5 different models (see Table S1) which are specific instantiations of the full model described above:

1. Model M_1 is the simplest model and contains only 2 parameters: $\theta = (a, \rho)$. It is assumed that $b = 1$ and that there is no refractoriness ($\gamma_k = 0$).

2. Model M_2 is the same as model 1 except that the *touch* priority distribution has 2 free parameters: a and b. It contains 3 parameters: $\theta = (a, b, \rho)$.

3. Model M_3 includes refractoriness (i.e. $\gamma_k \neq 0$, $k = 1, \ldots, 8$) but assumes $b = 1$. It contains therefore 10 parameters: $\theta = (a, \rho, \gamma_1, \ldots, \gamma_8)$.

4. Model M_4 is the same as model 3 except that the *touch* priority distribution is described by both a and b. It contains 11 parameters: $\theta = (a, b, \rho, \gamma_1, \ldots, \gamma_8)$.

5. Model M_5 is the same as model 4 except that there are $n = 12$ basis functions for the refractory kernel with $\alpha_1^{-1} = 50$ ms and $\alpha_{12}^{-1} = 1'500$ ms. It contains in total 15 parameters: $\theta = (a, b, \rho, \gamma_1, \ldots, \gamma_{12})$.

Model Fitting. For each model and for each subject, the model parameters θ are fitted from the set $\mathcal{D} = \{\tau_i\}_{i=1}^N$ of inter-touch intervals $\tau_i = t_i - t_{i-1}$. In order to do so, we relied on the continuous-time model which massively simplifies the expression of the log-likelihood. Indeed, the detailed model can be seen as a dynamical latent variable model (where the latent variables are x and y) which can be fitted through EM type algorithm but is known to be very slow. Here, because of the analytical expression of the ITI for the continuous-time model (see Eq. 13), we can express the following objective function

$$\mathcal{L}(\theta) = L(\theta) - \lambda \sum_{k=1}^n \gamma_k^2, \tag{15}$$

which is the log-likelihood $L(\theta) = \sum_{i=1}^N \log p(\tau_i)$ (see Eq. 20 in SI) minus a regularisation term on the coefficients γ_k to prevent overfitting. This regularisation term is only used in models 3-5 and $\lambda = 0.01$. Note that this objective function can be seen as the log-posterior with a Gaussian prior (with variance $1/2\lambda$) on the coefficients γ_k and a flat prior for the other parameters.

Because of the refractory kernel must remain positive for all time, i.t. $r(\tau) \geq 0$, $\forall \tau \geq 0$, the optimisation task can be expressed as

$$\theta^* = \arg\max_\theta \mathcal{L}(\theta) \quad \text{s.t.} \quad \sum_{k=1}^n \exp(-\alpha_k \tau)\gamma_k^* \geq -1 \quad \forall \tau \geq 0, \tag{16}$$

which can be treated as constrained optimisation problem with inequality constraints (see SI).

Fitting results. We recorded smartphone touches from 84 individuals for an average duration of 36.5 days (see SI for details on data collection). The average number of smartphone screen touches per day ranged from 285 to 9'915 with a median value of 2'540 touches per day.

For each individual, the 5 different models were fitted according to the procedure described above. By performing Bayesian model comparison (see SI), we found that model 3 explains best the data. Fig. 3 displays the results of model 3. We found that for each individual the empirical ITI distribution (see Fig. 3a1) is well captured by the model both for the short time scales (up to 1s) which is strongly influenced by the refractory kernel $r(t)$ (see Fig. 3b1) as well as the longer ITI which has a typical power-law decay. Note that because of the richness of the data, the power-law relationship extends over 5 decades (from 10^3 to 10^8 ms).

The fitted refractory kernel (see Fig. 3b1) shows a strong reduction of touching rate during the first few hundreds of milliseconds after the last touch and even displays a small increase in touching rate about 1s after the last touch. This smooth transition from short ITI to longer ITI removes the need to define an arbitrary onset of the power-law distribution [6].

The fitted touch priority distribution (see Fig. 3c1) (assuming that the *other* priority distribution is given by $q(y) = 1$) is concentrated around small priorities. Note that for model 3, we have $b = 1$ and therefore the *touch* priority distribution is given by $p(x) = ax^{a-1}$. For this class of distributions, if the mean *touch* priority $\langle x \rangle$ is smaller than the mean *other* priority $\langle y \rangle$ (which is the case in Fig. 3c1), then this relationship remains invariant under any priority transformation ϕ defined in Eq. (11). We can therefore state that for this subject and for model 3, interacting with the smartphone has on average a lower priority than doing other tasks.

We repeated this fitting procedure for the 84 subjects. The population results are displayed on Fig. 3a2-c2. We found that over the population the priority parameter a is fairly scattered around a median value of $a = 0.61$. Only 3 subjects have a priority parameter $a > 1$ indicating that for those subjects the average *touch* priority is larger than the average *other* priority. The large inter-individual differences is also highlighted in Fig. 3e which displays a broad distribution of touching rate ρ over the population. Note that our model is not restricted to rational power-law exponents. Indeed in our framework the power-law exponent is given by $a + 1$ where a can take any real positive value. In contrast, in the work of [4], the exponent is determined by the length of the list of tasks[8]. In our study, the (mean) power-law exponent is 1.63 ± 0.14 which is different from frequently found exponent of 1, 1.5 or 2.

Discussion. In this letter, we proposed a generalised priority-based model which is both flexible and tractable. The flexibility comes from the set of basis functions which describe refractory effects at short inter-touch intervals while the tractability stems from the simplified structure of the generative model in continuous-time which enables a fast fitting procedure. The flexibility is essential to capture inter-individual differences in touching behavior while the tractability is crucial for fitting large data sets.

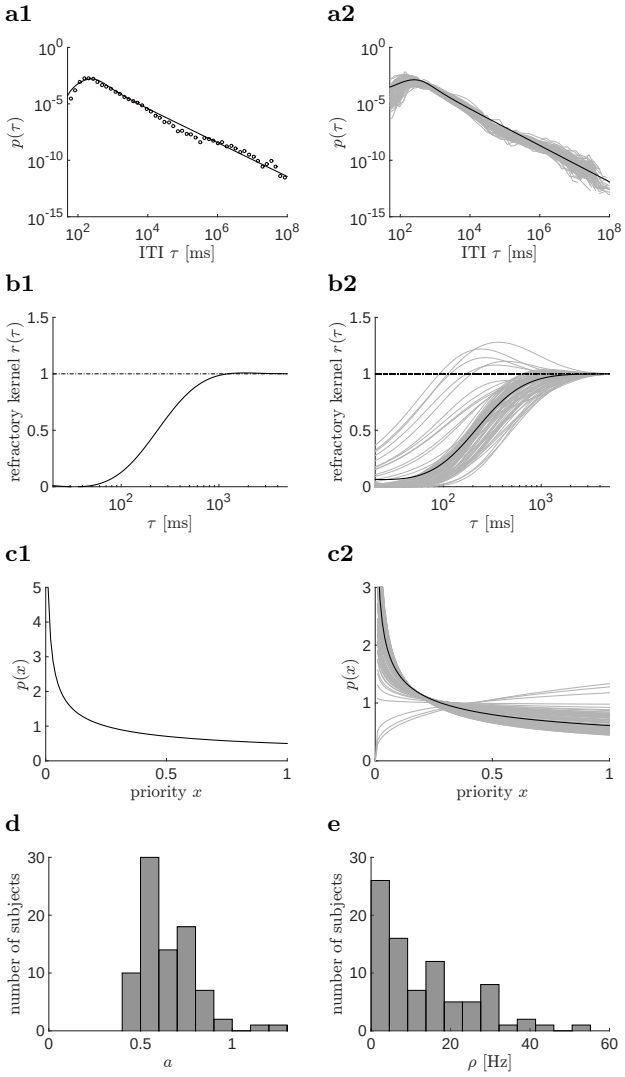

FIG. 3. Fitting results for one subject (**a1-c1**) and for the population of 84 subjects (**a2-c2**). **a1** The ITI distribution for one given subject (open circles) is well captured model (solid line). **b1** Refractory kernel. **c1** *touch* priority distribution (with $q = 1$). **a2-c2** Same as in **a1-c1** but for each of the 84 subjects (gray lines). Solid lines are obtained with the median of the fitted parameters i.e. $a = 0.61$, $b = 1$ and $\rho = 9.3$ Hz. **d** Distribution of the parameter a. **e** Distribution of the touching rate ρ.

Here, this generalised priority-based model has been applied to smartphone touching data, but could be applied to other event-based data sets which display power-law property for large inter-event intervals such as surface mails, emails or even foraging patterns.

We thanks Simone Carlo Surace for very helpful discussions. JPP was supported by the Swiss National Science Foundation grant PP00P3_150637. AG was supported by the Society in Science the Branco Weiss Fellowship and a research grant from the Holcim Foundation.

[1] J. G. Oliveira and A.-L. Barabási, *Human dynamics: Darwin and Einstein correspondence patterns.*, Vol. 437 (Nature, 2005).

[2] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, Physical Review E **73**, 036127 (2006).

[3] A. L. Barabasi, Nature **435**, 207 (2005).

[4] J. G. Oliveira and A. Vazquez, Physica A **388**, 187 (2009).

[5] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. N. Amaral, Proceedings of the National Academy of Sciences **105**, 18153 (2008).

[6] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Review **51**, 661 (2009).

[7] P. Brémaud, *Point processes and queues*, Vol. 30 (Springer, 1981).

[8] Technically, [4] assume that an event occurs only if the interacting task for both agents A and B have the highest priority compared to all other tasks of length L_A (for agent A) and L_B for agent B. If $L = L_A = L_B$, then the power-law exponent is given by $1 + 1/(L - 1)$.

[9] M. Balerna and A. Ghosh, npj, Digital Medicine, in Press (2018).

SUPPLEMENTARY INFORMATION

Invariance of the model

In this section, we will show that that the ITI distribution remains unchanged if the pair of priority distribution $(p(x), q(y))$ is replaced by $(\tilde{p}(x), \tilde{q}(y))$ where $\tilde{p}(x)$ and $\tilde{q}(y)$ are given by Eq. (11).

Let us consider the following change of variable: $x = \phi(x')$. The ITI distribution can be therefore expressed as

$$p(\tau) = \int_0^1 p^q(\tau|\phi(x'))p(\phi(x'))\phi'(x')dx', \tag{17}$$

where the conditional ITI distribution $p^q(\tau|\phi(x'))$ depends on the *other* priority distribution $q(y)$ via the instantaneous rate $\bar{\lambda}^q(\phi(x'), \tau)$ which can be expressed as

$$\begin{aligned}\bar{\lambda}^q(\phi(x'), \tau) &= \rho r(\tau) \int_0^{\phi(x')} q(y)dy \\ &= \rho r(\tau) \int_0^{x'} q(\phi(y))\phi'(y)dy = \bar{\lambda}^{\tilde{q}}(x', \tau),\end{aligned} \tag{18}$$

where $\tilde{q}$ is given by Eq. (11). Note that the dependence on q is included only here for the clarity of the argument, but is omitted otherwise for the simplicity of the notation. Therefore the ITI distribution is invariant under the change of variable ϕ for both x and y. Indeed, we have

$$\begin{aligned}p(\tau) &= \int_0^1 p^q(\tau|x)p(x)dx \\ &= \int_0^1 p^{\tilde{q}}(\tau|x)\tilde{p}(x)dx\end{aligned} \tag{19}$$

For example, if the *touch* priority distribution is given by $p(x) = \text{Beta}(x; a, 1)$ and the *other* priority distribution is given by $q(y) = \text{Beta}(y; a', 1)$, then the function $\phi(x) = x^k$ allows to generate a family of equivalent pairs of priority distributions $(\tilde{p}(x), \tilde{q}(y)) = (\text{Beta}(x; ka, 1), \text{Beta}(y; ka', 1))$. Therefore, the ITI remains unchanged as long as a/a' remains constant.

Log-likelihood gradient

We fitted the parameters $\theta = (a, b, c, \gamma_1, \ldots, \gamma_n)$ by performing maximum likelihood with a suitable regularisation for the parameters γ_i. Note that for a practical implementation, it is easier to learn $c = \log(\rho)$ instead of ρ itself. For a set of inter-touch intervals $\mathcal{D} = \{\tau_i\}_{i=1}^N$, the log-likelihood can be expressed as

$$L(\theta) = Nc + \sum_{i=1}^N \log(r(\tau_i)) + \log\left(\langle xE_i(x)\rangle\right), \tag{20}$$

where the expectation $\langle\cdot\rangle$ is w.r.t $p(x) = \text{Beta}(x; a, b)$ and the function $E_i(x)$ is given by

$$E_i(x) = \mathrm{e}^{-\rho x R(\tau_i)}, \tag{21}$$

and $R(\tau_i)$ is given by

$$R(\tau_i) := \int_0^{\tau_i} r(t)dt = \tau_i + \sum_{k=1}^n \frac{\gamma_k}{\alpha_k}\left(1 - \mathrm{e}^{-\alpha_k \tau_i}\right). \tag{22}$$

By noting that

$$\frac{\partial \log(p(x))}{\partial a} = \log(x) - \langle \log(x)\rangle, \tag{23}$$

we can compute the log-likelihood gradient w.r.t a:

$$\frac{\partial L}{\partial a} = \sum_{i=1}^N \frac{\text{cov}(xE_i(x), \log(x))}{\langle xE_i(x)\rangle}. \tag{24}$$

By symmetry, the gradient of L w.r.t to b yields

$$\frac{\partial L}{\partial b} = \sum_{i=1}^N \frac{\text{cov}(xE_i(x), \log(1-x))}{\langle xE_i(x)\rangle}. \tag{25}$$

The gradient of L w.r.t c is given by

$$\frac{\partial L}{\partial c} = N - \rho \sum_{i=1}^N \frac{\langle x^2 E_i(x)\rangle}{\langle xE_i(x)\rangle} R(\tau_i), \tag{26}$$

Finally, the gradient of L w.r.t γ_k can be expressed as

$$\begin{aligned}\frac{\partial L}{\partial \gamma_k} &= \sum_{i=1}^N \frac{\partial r(\tau_i)/\partial \gamma_k}{r(\tau_i)} - \rho \frac{\langle x^2 E_i(x)\rangle}{\langle xE_i(x)\rangle} \frac{\partial R(\tau_i)}{\partial \gamma_k} \\ &= \sum_{i=1}^N \frac{\mathrm{e}^{-\alpha_k \tau_i}}{r(\tau_i)} - \rho \frac{\langle x^2 E_i(x)\rangle}{\langle xE_i(x)\rangle} \frac{(1 - \mathrm{e}^{-\alpha_k \tau_i})}{\alpha_k}\end{aligned} \tag{27}$$

Computing the integrals

Both the log-likelihood L as well its gradient w.r.t to the parameters θ contain integrals that are delicate to evaluate. Indeed, the integrand of all those integrals depend on the Beta distribution $\text{Beta}(x; a, b)$ which can diverge at $x = 0$ or $x = 1$ depending on the parameters a and b. So whenever possible, we compute those integrals analytically. This can be done for the following integrals

$$\langle \log(x)\rangle_{a,b} = \frac{d}{da} B(a, b) = \psi(a) - \psi(a + b), \tag{28}$$

where $\psi(z) = d \log \Gamma(z)/dz$ is the digamma function and $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ is the Beta function. By symmetry, we have

$$\langle \log(1 - x)\rangle_{a,b} = \psi(b) - \psi(a + b). \tag{29}$$

By Taylor expanding the exponential in the expression of $E_i(x)$, the integral $\langle xE_i(x)\rangle_{a,b}$ can be expressed as

$$\begin{aligned}\langle xE_i(x)\rangle_{a,b} &= \frac{a}{a+b} \langle E_i(x)\rangle_{a+1,b} \\ &= \frac{a}{a+b} \, {}_1F_1(a + 1, a + b + 1; -\rho R(\tau_i))\end{aligned} \tag{30}$$

where ${}_1F_1$ is the hypergeometric function defined as

$$ {}_1F_1(a,b;z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} \frac{(a)_k}{(b)_k} \tag{31}$$

and $(a)_k = \prod_{i=0}^{k-1}(a+k)$ for $k \geq 1$ (and $(a)_0 = 1$) is the rising factorial (also called Pochhammer function). Similarly, $\langle x^2 E_i(x) \rangle_{a,b}$ can be expressed as

$$ \langle x^2 E_i(x) \rangle_{a,b} = \frac{B(a+2,b)}{B(a,b)} \, {}_1F_1(a+2, a+b+2; -\rho R(\tau_i)). \tag{32}$$

When it is not possible to compute the integrals analytically, the idea is to express the integral into a sum of two integrals where the first one is well suited for a numerical integration and the second one can be performed analytically. For example $\langle x E_i(x) \log(1-x) \rangle_{a,b}$ can be computed as

$$\begin{aligned} &\langle x E_i(x) \log(1-x) \rangle_{a,b} \\ &= \frac{a}{a+b} \Big\{ \langle (E_i(x) - E_1(x) \log(1-x)) \rangle_{a+1,b} \\ &+ E_i(1) \langle \log(1-x) \rangle_{a+1,b} \Big\}, \end{aligned} \tag{33}$$

where the first term of the r.h.s can be computed numerically and the second term can be computed with Eq. (29).

Finally, It should be noted that the integral $\langle x E_i(x) \log(x) \rangle_{a,b}$ can be computed numerically straightforwardly since the integrand does not diverges when $x = 0$ nor when $x = 1$.

Constrained optimisation

The difficulty of the optimisation problem defined in Eq. (16) lies in the fact that the constraints are defined for all $\tau \geq 0$ (i.e. infinitely many inequality constraints). For a practical numerical implementation, we defined a grid of $M = 100$ points $\tau_1, \ldots, \tau_M$ and imposed that the constraints are satisfied on those points, i.e.

$$ \theta^* = \arg\max_{\theta} \mathcal{L}(\theta) \quad \text{s.t.} \quad \sum_{k=1}^{n} \exp(-\alpha_k \tau_i) \gamma_k^* \geq -1 \tag{34}$$

$\forall i = 1, \ldots, M$. The parameters θ are learned through gradient ascent of the objective function $\mathcal{L}(\theta)$. Whenever the proposed parameter θ' given by

$$ \theta' = \theta + \eta \nabla_\theta \mathcal{L}(\theta) \tag{35}$$

remains inside the allowed set of parameters Θ, then the proposed parameter is accepted, i.e. $\theta \leftarrow \theta'$. If the proposed parameter $\theta' \notin \Theta$, say because the constraint is not satisfied on the i^{th} time point (i.e. $r(\tau_i) \geq 0$ is violated) then it is projected back on the boundary $\partial\Theta$, i.e.

$$ \theta \leftarrow \theta' - \frac{1 + \left(\mathbf{w}^{(i)}\right)^T \theta'}{\|\mathbf{w}^{(i)}\|^2} \mathbf{w}^{(i)}, \tag{36}$$

where $\mathbf{w}^{(i)} = (0, 0, 0, \exp(-\alpha_1 \tau_i), \ldots, \exp(-\alpha_n \tau_i))^T$ is the vector perpendicular to the hyperplane defined by $\mathbf{w}^T \theta = -1$ and is valid for models 4 and 5. Note that for model 3, $\mathbf{w}$ is given by $\mathbf{w}^{(i)} = (0, 0, \exp(-\alpha_1 \tau_i), \ldots, \exp(-\alpha_n \tau_i))^T$ since there are only 2 non-gamma parameters (a and ρ). For model 1 and 2, a simple gradient ascent can be followed.

Model comparison

To compare the different models (see Table S1), we used the Bayesian information criterion which is given by $BIC = \log(N)|\theta| - 2\mathcal{L}(\theta^*)$ where N is the number of data points, $|\theta|$ is the number of parameters and $\mathcal{L}(\theta^*)$ is the objective function given by Eq. (15) and is evaluated at the MAP parameter θ^*. We found that for 82 (out of 84) subjects $\Delta BIC_{2,1} = BIC(M_2) - BIC(M_1) < -10$ thereby favoring M_2 over M_1 (see Fig. S1). Similarly, we found that according to the same criterion, M3 is favored over M2 for all subjects, M3 is favored over M4 for 74 subjects and M3 is favored over M5 for 72 subjects. So overall model 3 is favored over the other models.

Model	# of parameters	parameters
M_1	2	$\theta = (a, \rho)$
M_2	3	$\theta = (a, b, \rho)$
M_3	10	$\theta = (a, \rho, \gamma_1, \ldots, \gamma_8)$
M_4	11	$\theta = (a, b, \rho, \gamma_1, \ldots, \gamma_8)$
M_5	15	$\theta = (a, b, \rho, \gamma_1, \ldots, \gamma_{12})$

TABLE S1. List of models.

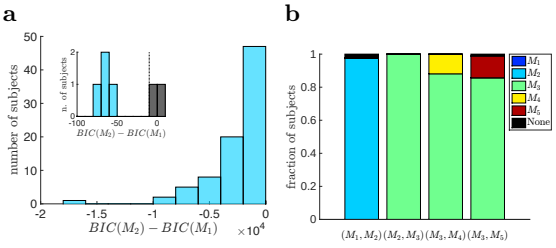

FIG. S1. Model comparison. **a** Distribution of the difference between the Bayesian Information Criterion (BIC) between model M_2 and model M_1. For 82 out of 84 subjects, model M_2 is preferred over model M_1 ($\Delta BIC < -10$). **b** For each pair of models (M_i, M_j), the fraction of subjects favoring model M_i is computed by the fraction of subject which obey $BIC(M_i) - BIC(M_j) < -10$. If $|BIC(M_i) - BIC(M_j)| < 10$ no model is favored (black). Overall, model M_3 is favored over the other ones.

Smartphone data collection

A custom-designed software application (app, Touchometer) that could record the touchscreen events with a maximum error of 5 ms [9] was installed on each participant's phone. To determine this accuracy, controlled test touches were done at precisely 150, 300 and 600 ms while the Touchometer recorded at 147, 301 and 600 ms respectively, with standard deviations less than 15 ms (interquartile range less than 5 ms). The app posed as a service to gather the timestamps of touchscreen events that were generated when the screen was in an unlocked state. The operation was verified in a subset of phones by using visually monitored tactile events. The data was stored locally and transmitted by the user at the end of the study via secure email. One subject was eliminated as the app intermittently crashed after a software update. The smartphone data were processed by using MATLAB (MathWorks, USA).